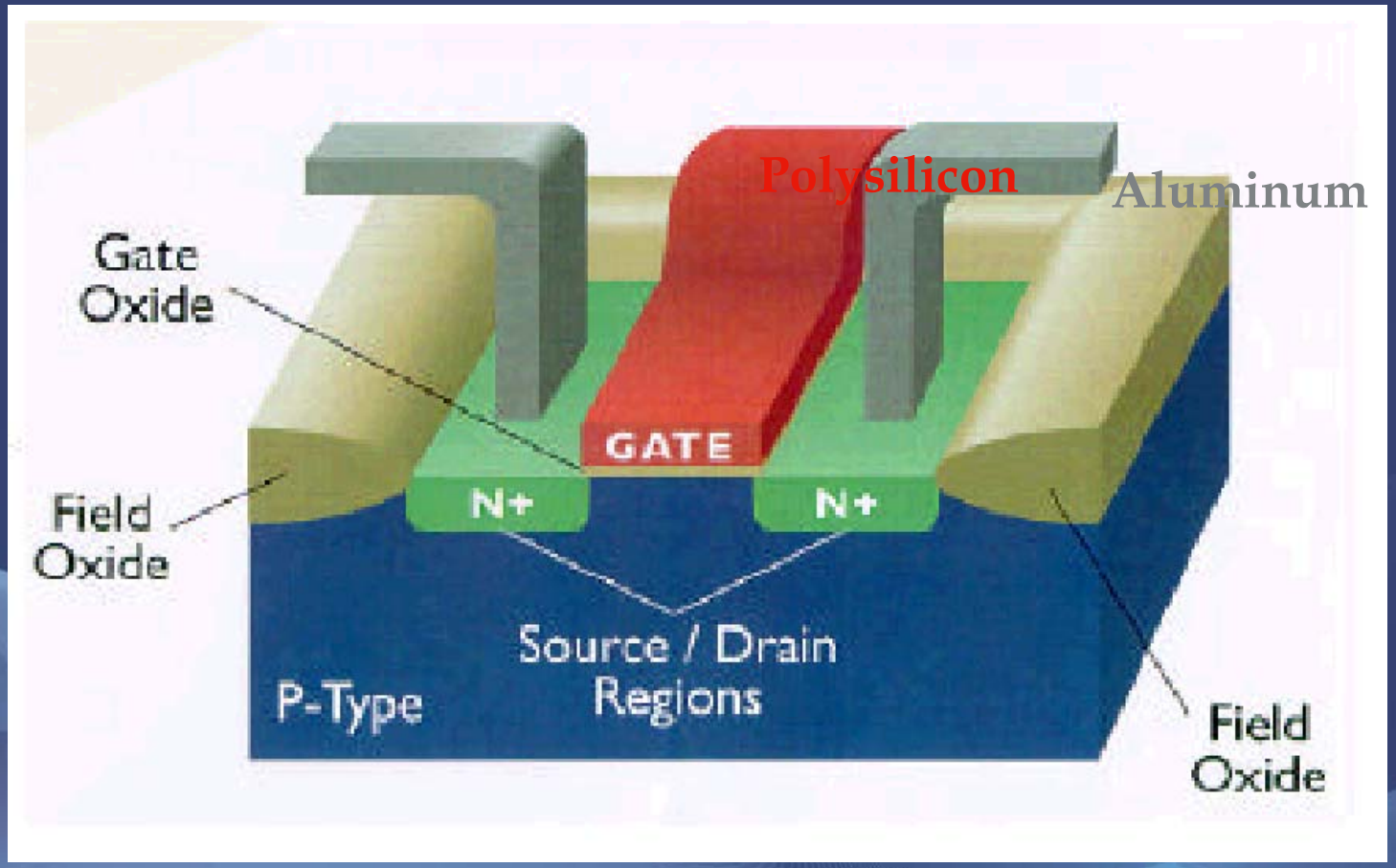
ECE 225
High-Speed Digital IC Design
Lecture 10

MOSFET Scaling

Prof. Kaustav Banerjee
Electrical and Computer Engineering
E-mail: kaustav@ece.ucsb.edu

The MOS Transistor



Review Lecture 7 and 8 from ECE 124A for basics of MOSFET Operation…

Goals of Technology Scaling

- Making things cheaper
  - Integrating more transistors per chip for the same money
  - Build same products cheaper, sell same part for less money
  - Price of fabrication of a transistor to be reduced
- But also want the transistor to be faster, smaller and low power

Technology Scaling

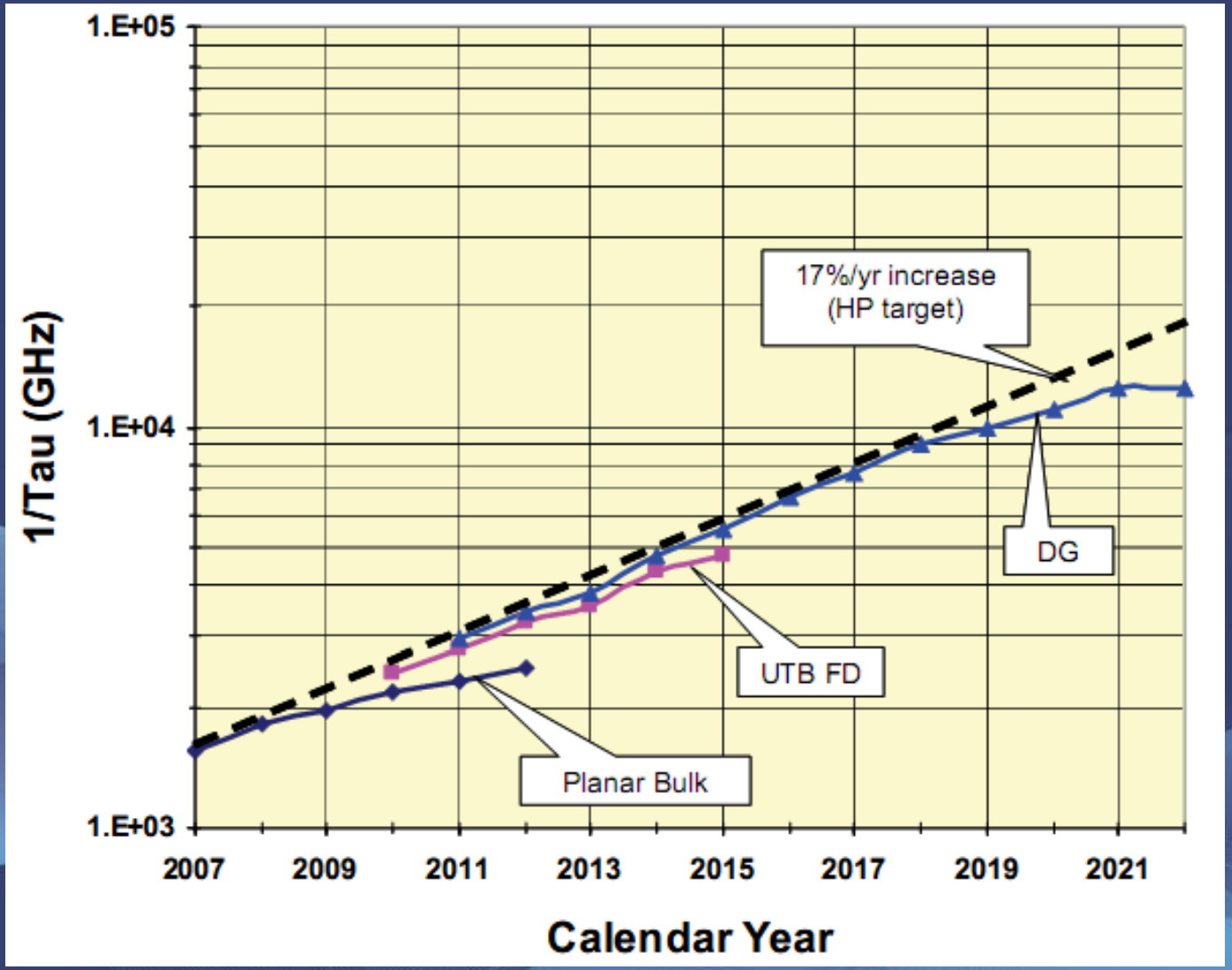
- Technology generation spans 2-3 years
- ☐Benefits of scaling the dimensions by 30%:
  - ☐Reduce gate delay by 30% (increase operating frequency by 43%)
  - ☐Double transistor density
  - ☐Reduce energy per transition by 65% (50% power savings @ 43% increase in frequency
- ☐Die size used to increase by 14% per generation

Technology Roadmap

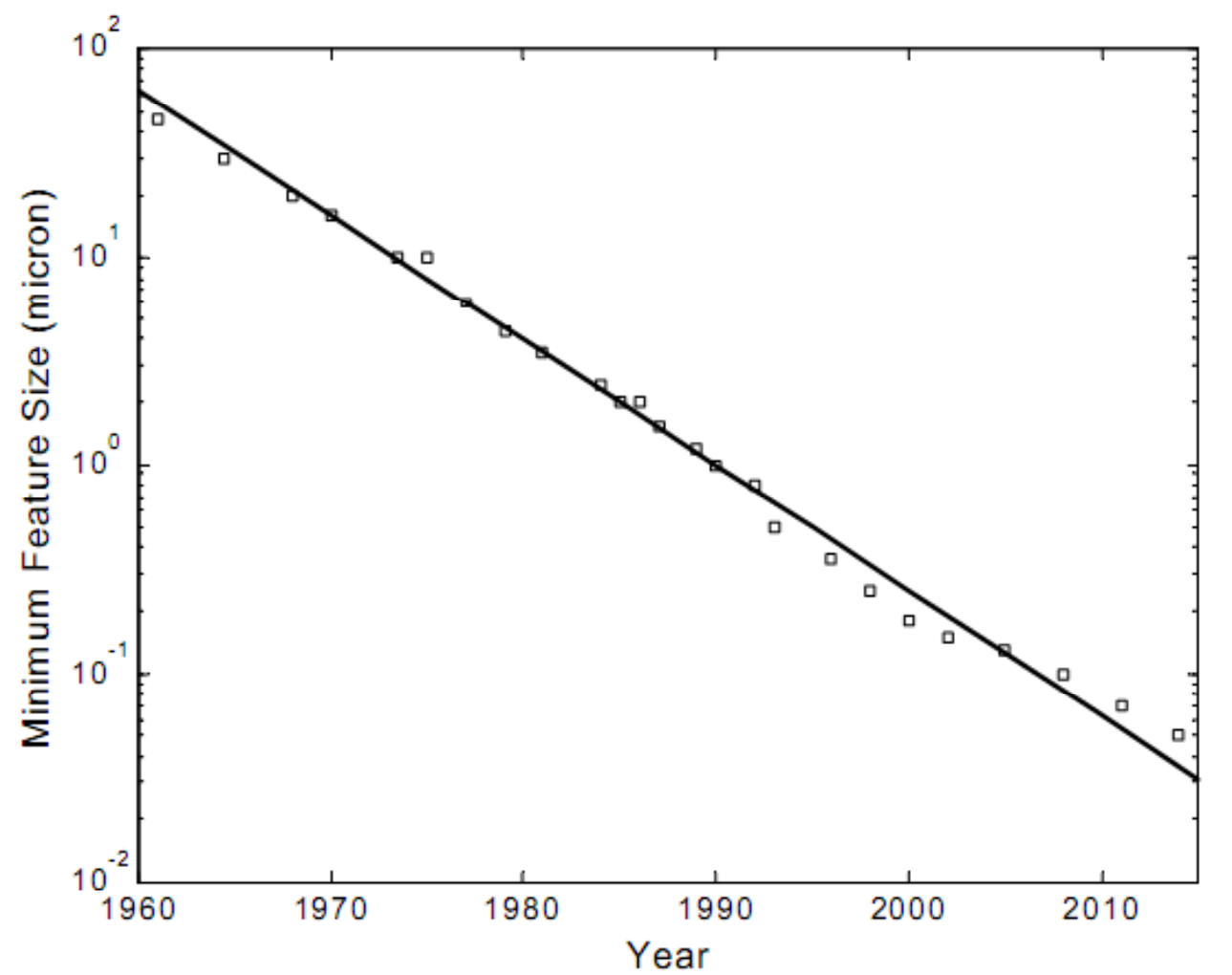
- International Technology Roadmap for Semiconductors 2002 data

Year	2001	2003	2005	2007	2010	2013	2016
DRAM ½ pitch [nm]	130	100	80	65	45	32	22
MPU transistors/chip	97M	153M	243M	386M	773M	1.55G	3.09G
Wiring levels	8	8	10	10	10	11	11
High-perf. phys. gate [nm]	65	45	32	25	18	13	9
High-perf. VDD [V]	1.2	1.0	0.9	0.7	0.6	0.5	0.4
Local clock [GHz]	1.7	3.1	5.2	6.7	11.5	19.3	28.8
High-perf. power [W]	130	150	170	190	218	251	288
Low-power phys. gate [nm]	90	65	45	32	22	16	11
Low-power VDD [V]	1.2	1.1	1.0	0.9	0.8	0.7	0.6
Low-power power [W]	2.4	2.8	3.2	3.5	3.0	3.0	3.0

Acceleration Continues

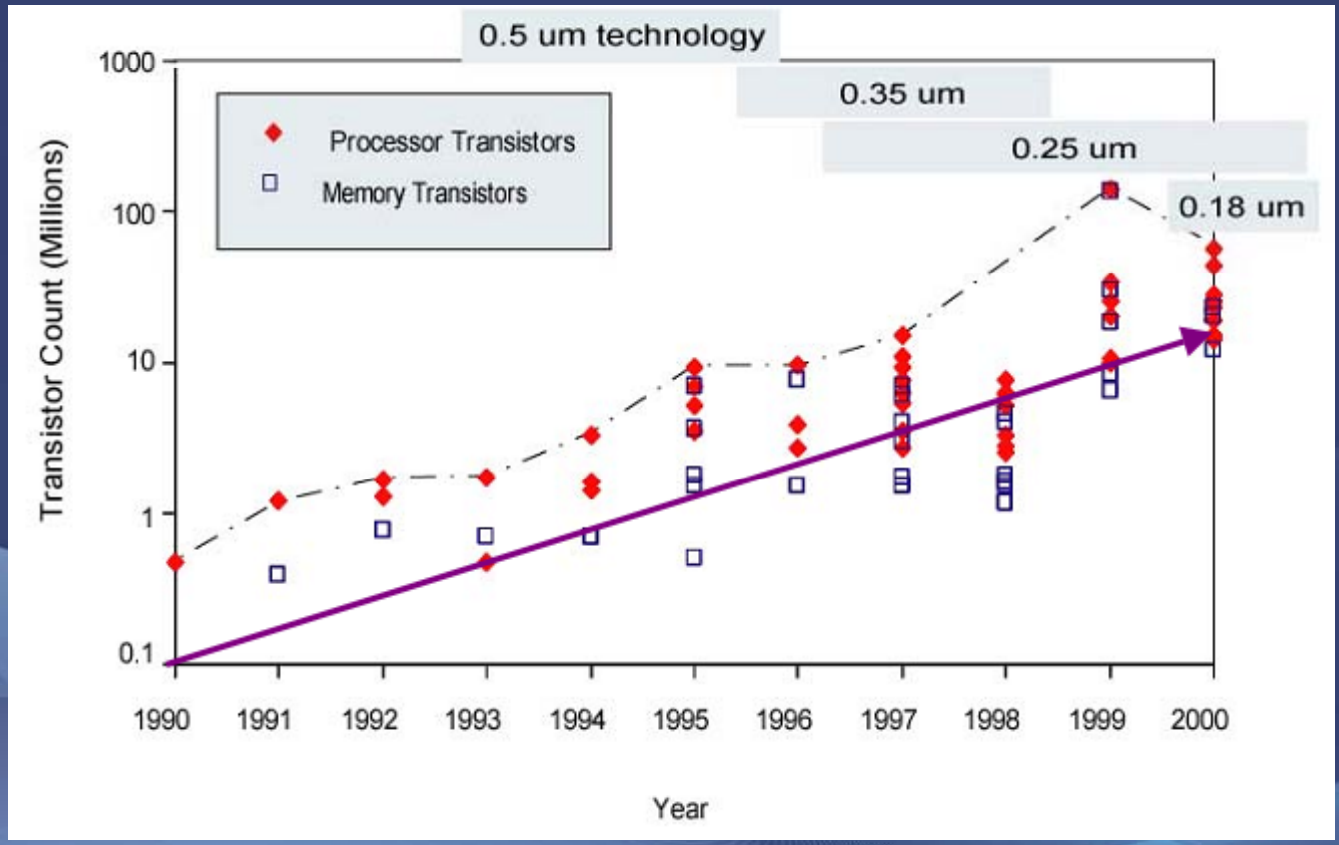


Technology Scaling



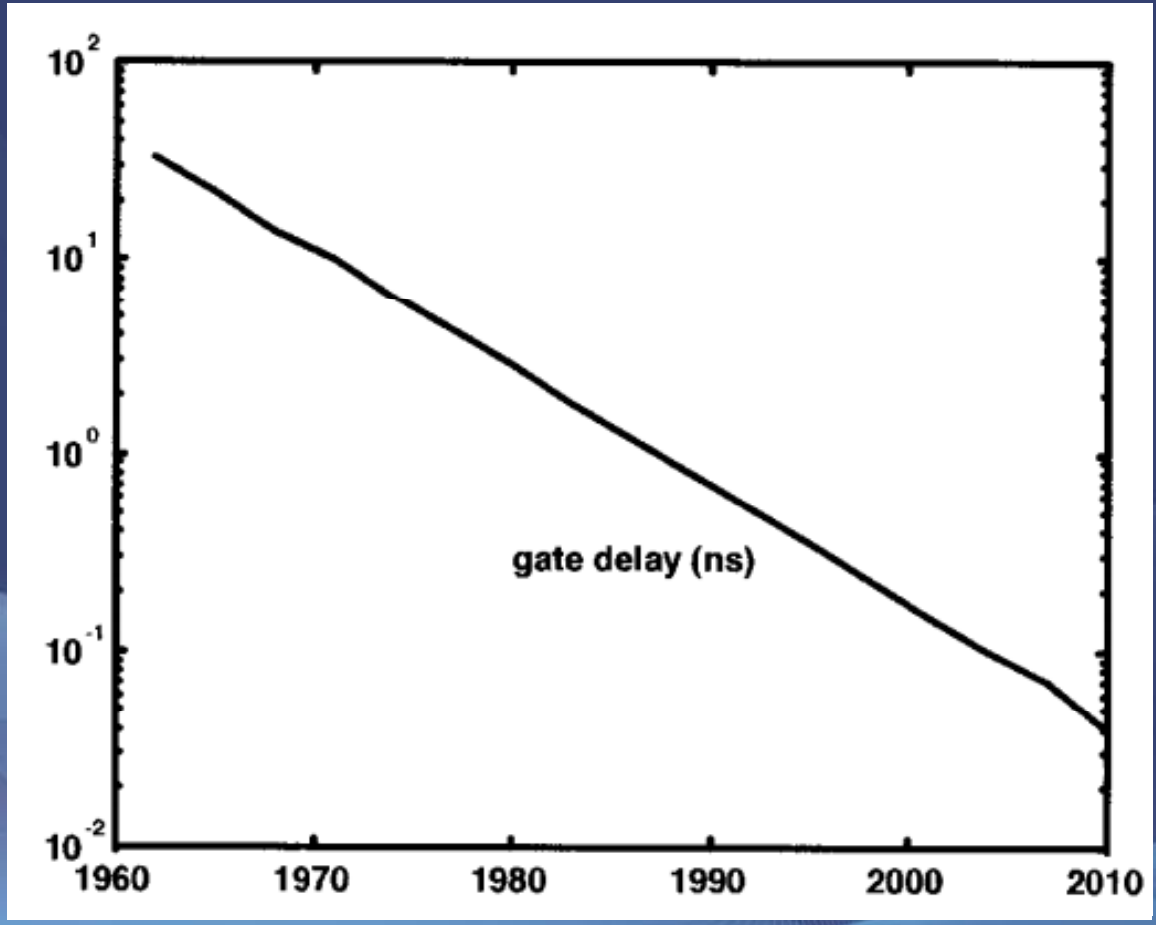
Minimum Feature Size

Technology Scaling



Number of components per chip

Technology Scaling



Propagation Delay

Technology Scaling Models

- Full Scaling (Constant Electrical Field)
  - ideal model, dimensions and voltage scale together by the same factor S
- Fixed Voltage Scaling
  - most common model until recently
  - only dimensions scale, voltages remain constant
- General Scaling
  - most realistic for todays situation
  - voltages and dimensions scale with different factors

Scaling Relationships for Long Channel Devices

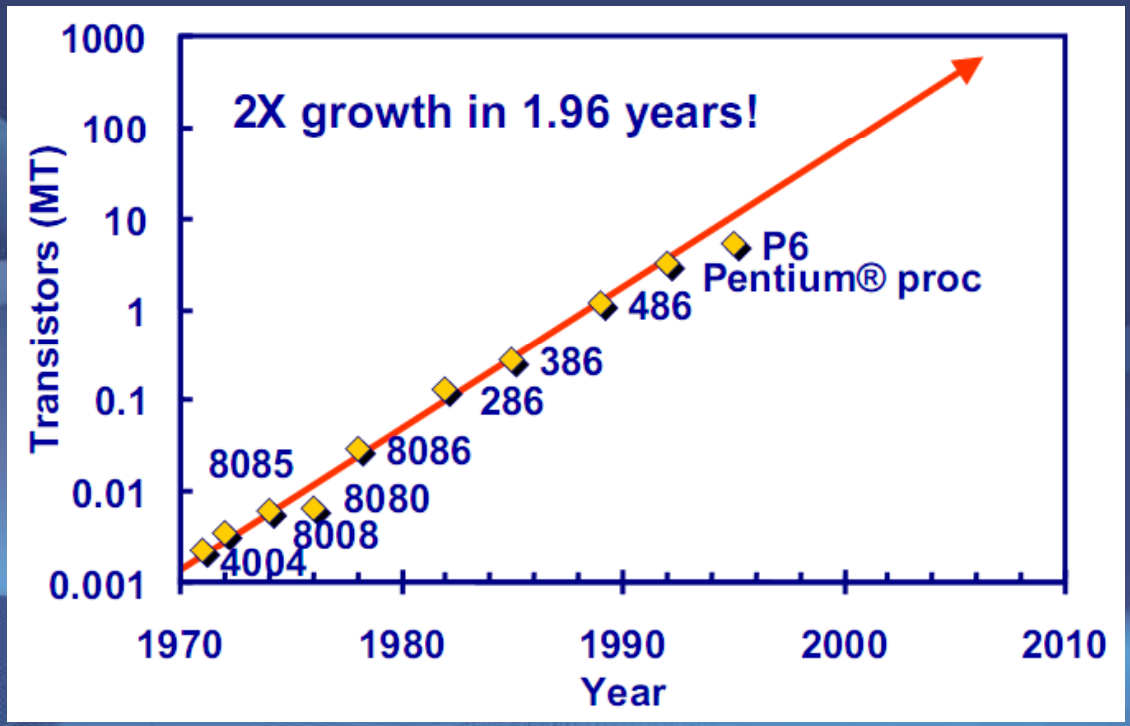
Parameter	Relation	Full Scaling	General Scaling	Fixed Voltage Scaling
W, L, t_{ox}		1/S	1/S	1/S
V_{DD}, V_T		1/S	1/U	1
N_{SUB}	V/W_{depl}^2	S	S^2/U	S^2
Area/Device	WL	$1/S^2$	$1/S^2$	$1/S^2$
C_{ox}	$1/t_{ox}$	S	S	S
C_L	$C_{ox}WL$	1/S	1/S	1/S
k_n, k_p	$C_{ox}W/L$	S	S	S
I_{av}	$k_{n,p} V^2$	1/S	S/U^2	S
t_p (intrinsic)	$C_L V / I_{av}$	1/S	U/S^2	$1/S^2$
P_{av}	$C_L V^2 / t_p$	$1/S^2$	S/U^3	S
PDP	$C_L V^2$	$1/S^3$	$1/SU^2$	1/S

Transistor Scaling for Velocity Saturated Devices

Parameter	Relation	Full Scaling	General Scaling	Fixed-Voltage Scaling
W, L, t_{ox}		$1/S$	$1/S$	$1/S$
V_{DD}, V_T		$1/S$	$1/U$	1
N_{SUB}	V/W_{depl}^2	S	S^2/U	S^2
Area/Device	WL	$1/S^2$	$1/S^2$	$1/S^2$
C_{ox}	$1/t_{ox}$	S	S	S
C_{gate}	$C_{ox}WL$	$1/S$	$1/S$	$1/S$
k_n, k_p	$C_{ox}W/L$	S	S	S
Current Density	I_{sat}/Area	S	S^2/U	S^2
Ron	V/I_{sat}	1	1	1
P	$I_{sat}V$	$1/S^2$	$1/U^2$	1

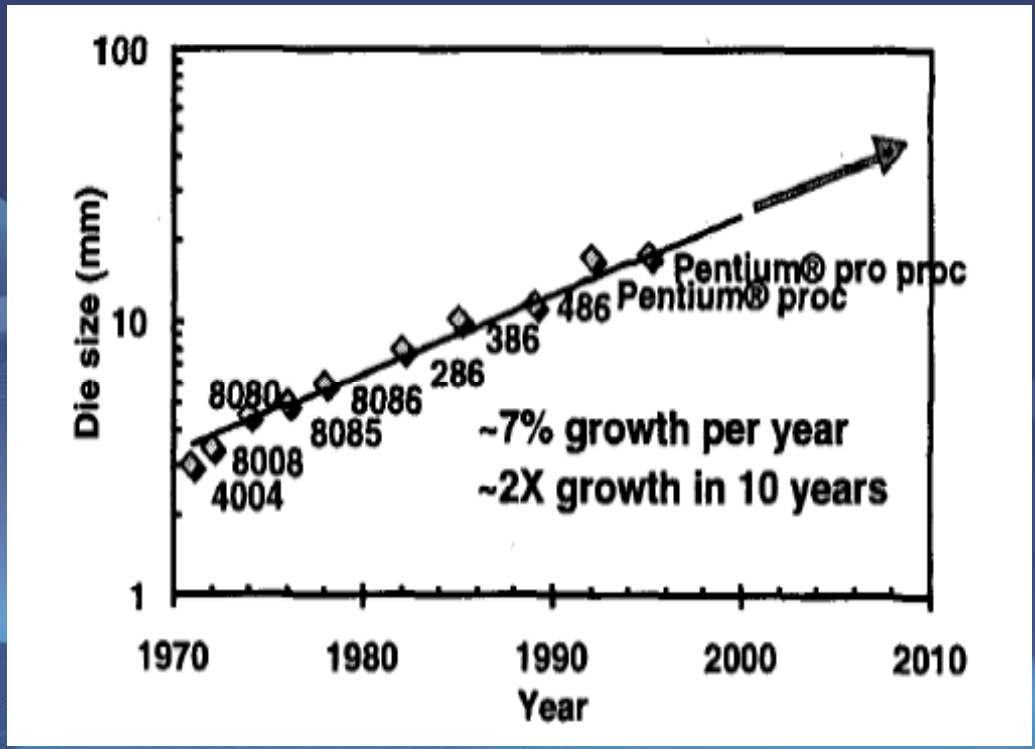
Microprocessor Scaling

- Processors used for comparison are 'Lead Microprocessors'.
- Indicates a doubled growth in transistor count every 2 years.



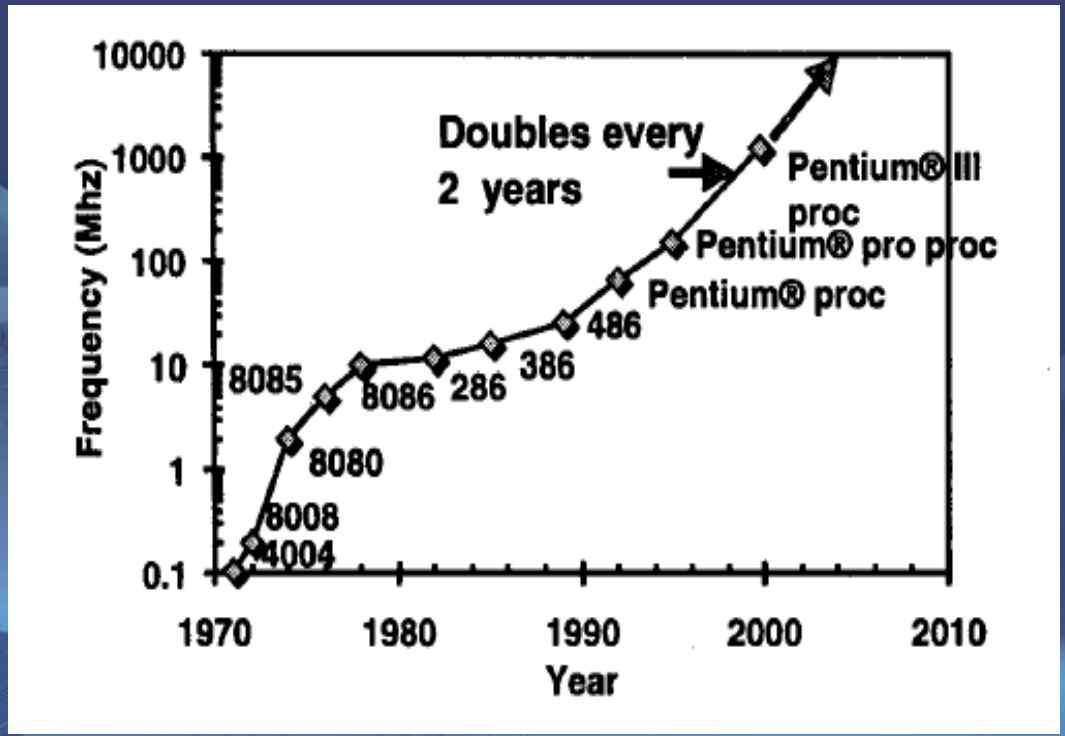
Die Size Trend

- Growth in die size with technology scaling.
- Graph indicates 7% growth per year and doubled growth in 10 years.



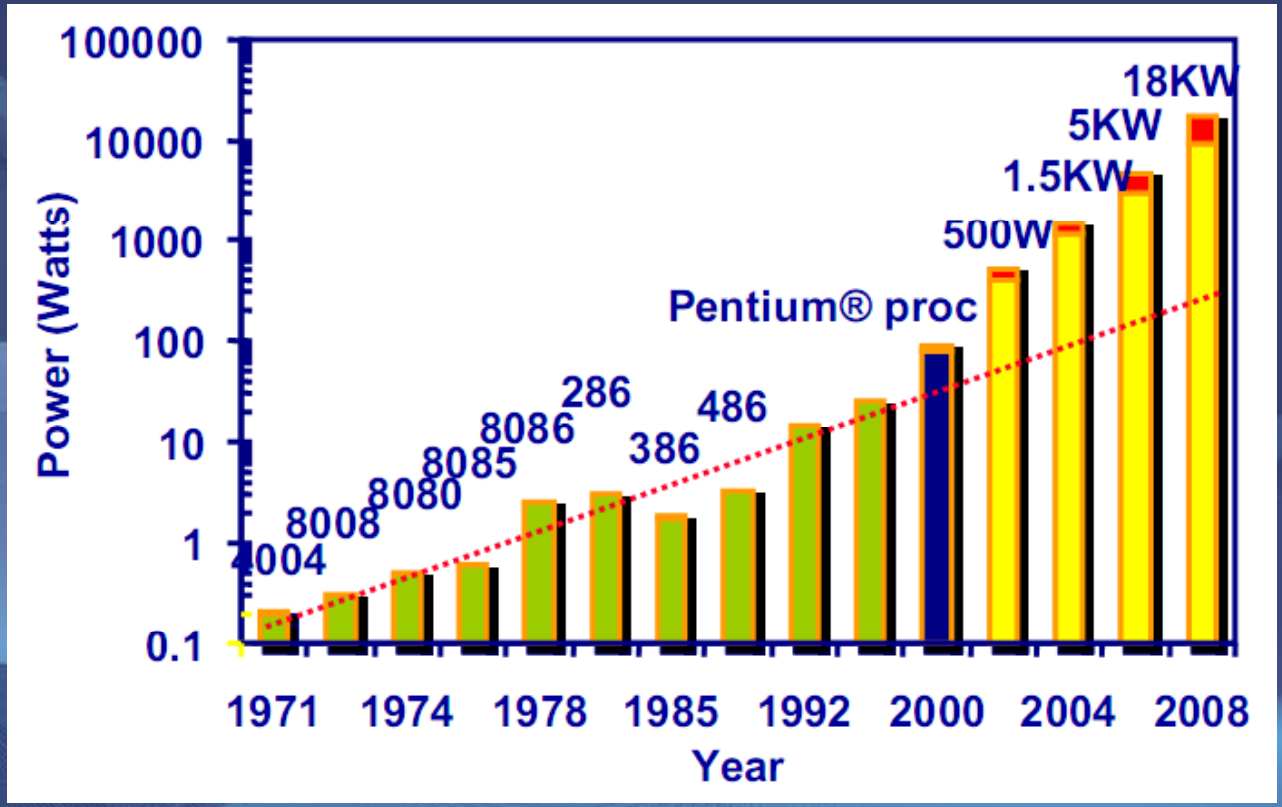
Operation Frequency Trend

- Initially increase in operating frequency was rapid.
- In the recent past operating frequency has been doubling every two years.

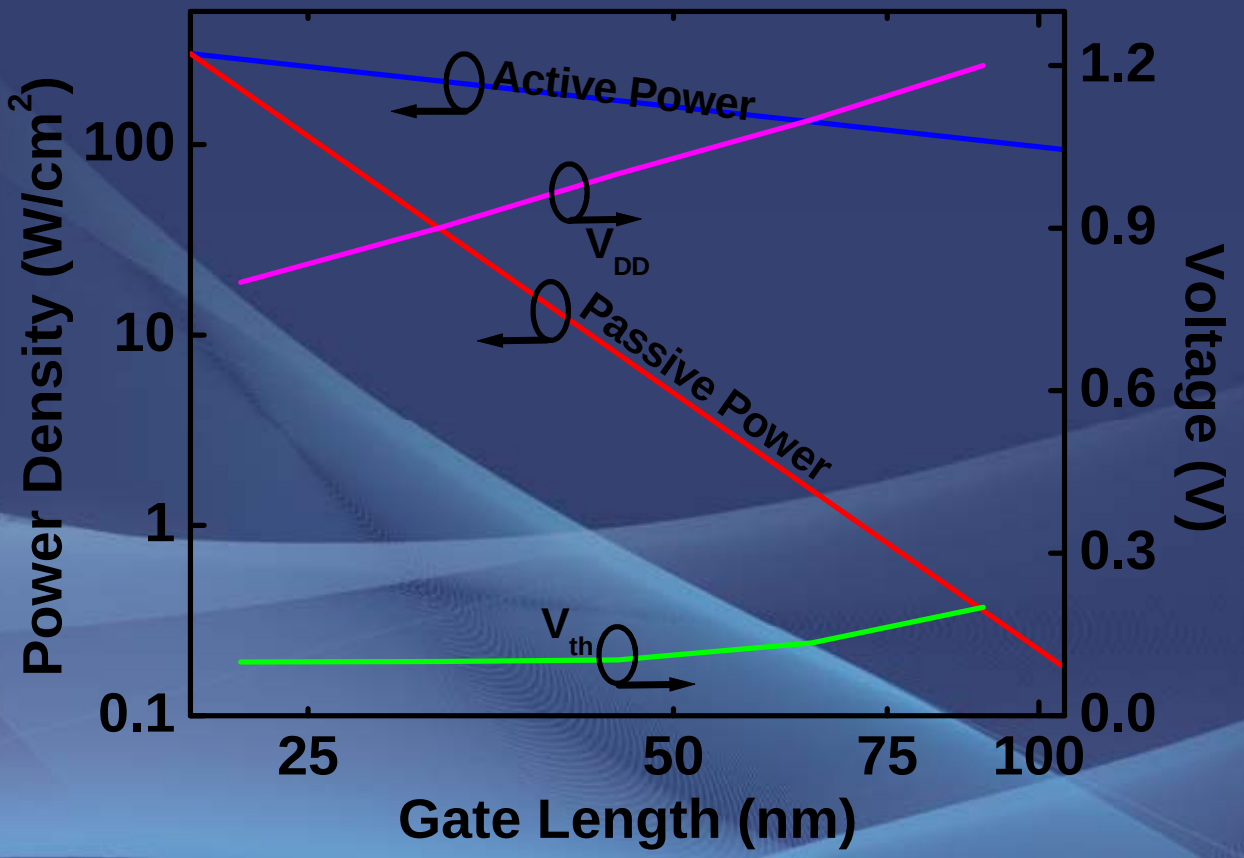


Microprocessor Power

- Continuous increase in power dissipation except for the one dip.
- Dip caused due to NMOS to CMOS transition.

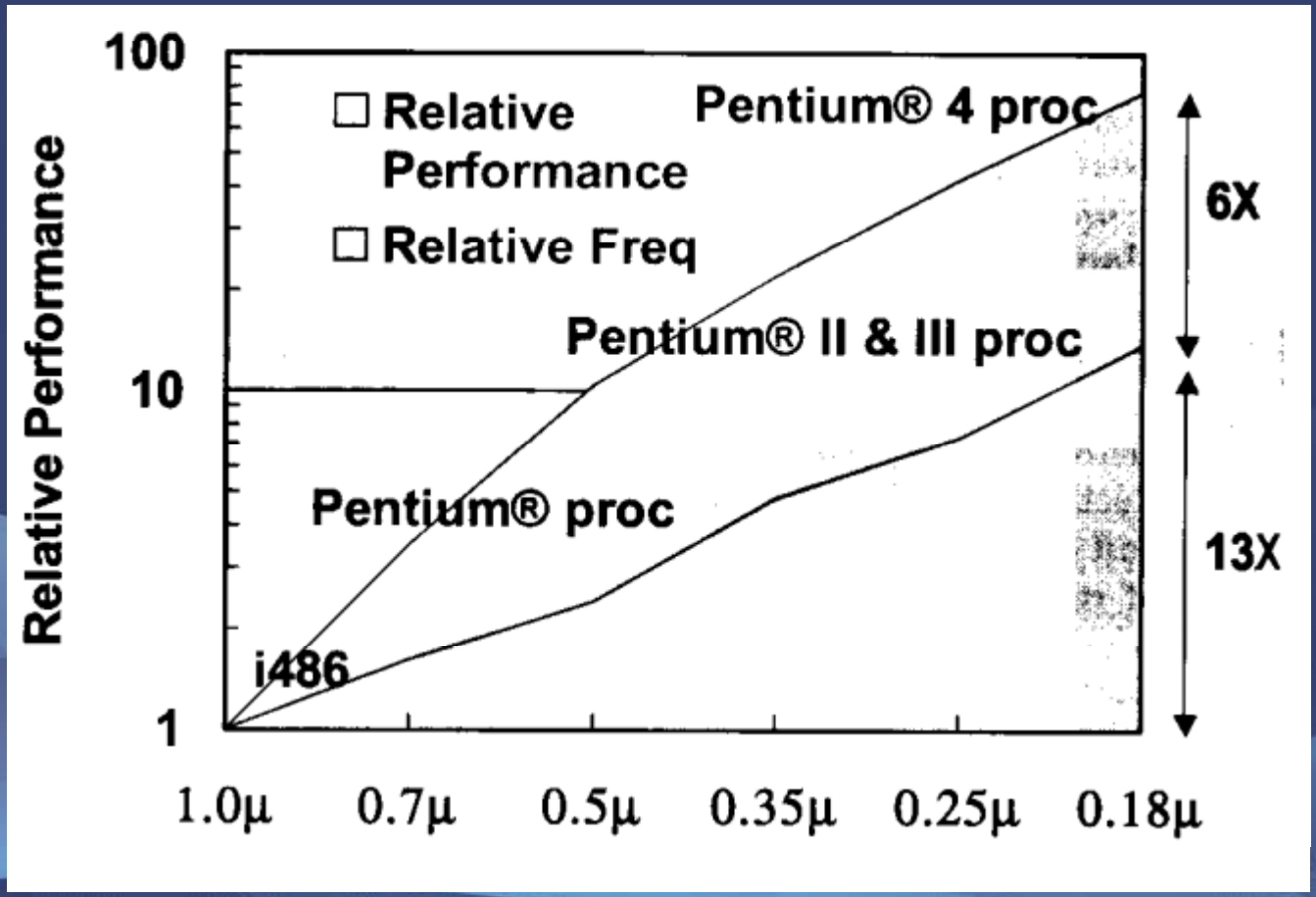


Gate Length Scaling Trends



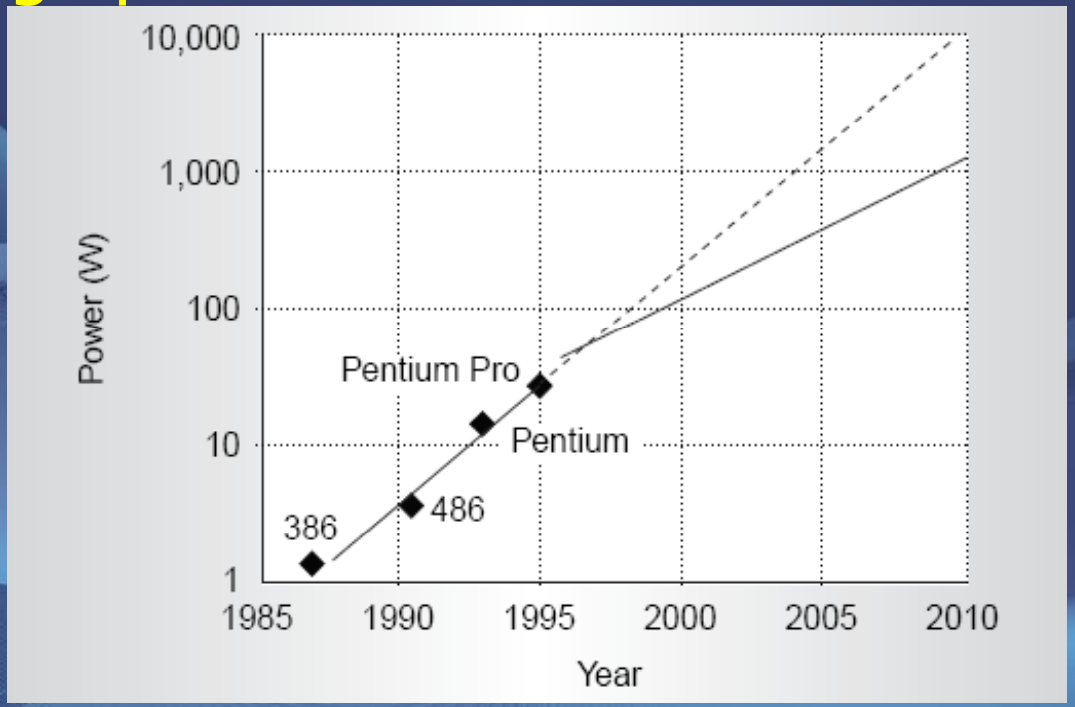
Puri *et. al., Nano Lett.,* 2003

Microprocessor Performance



Power Dissipation Projection

- Considering only active power, power dissipation increases from 100W in 1999 to 2000W in 2010.
- Dotted line indicates the power dissipation including the leakage power.



Summary of trends

- Assuming technology scaling continues following the trend seen before, the following characteristics are observed every technology generation (i.e. every 2 years).
  - Electrical nodes in a given area doubles.
  - Die size grows by 14%.
  - Supply voltage decreases.
  - Frequency doubles.

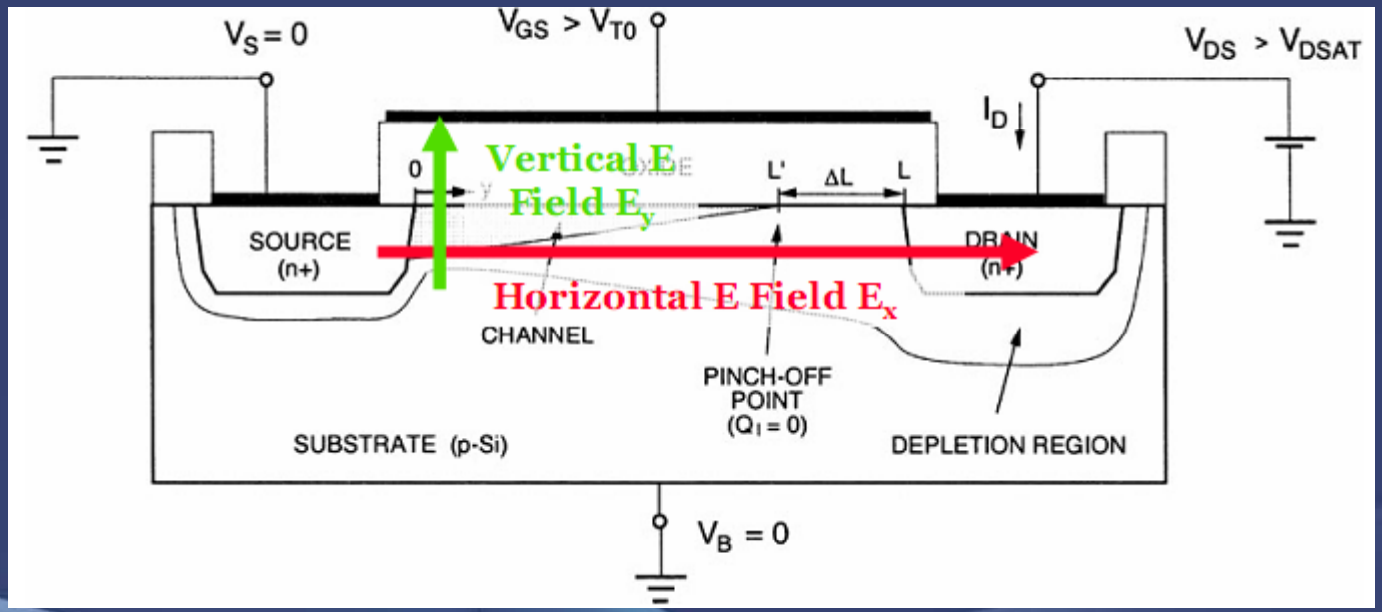
Problems Faced

- Threshold voltage to be scaled aggressively
- Reduced Vth results in higher leakage current
- Temperature on the die increases → higher resistance
- Interconnect delay becomes significant
- Soft error rate increases with scaling

Short Channel Effects

- Short-channel device: channel length is comparable to depth of drain and source junctions and depletion width
  - In general, visible when L ~ 1μm and below
- Short-channel effects
  - Carrier velocity saturation
  - Mobility degradation
  - Threshold voltage variation
    - VT Roll-Off
    - Drain-Induced Barrier Lowering
  - Oxide breakdown
  - Avalanche breakdown

Velocity Saturation



Vertical Electric Field

1980 E_Y=5V/1000A° =50 x 10^4V/cm

1995 E_Y=3.3.V/75A°=4.4 x 10^6V/cm

2001 E_y=1.2V/22AO=5.5 x 10^6 V/cm

Horizantal Electric Field

1980 E_x=5V/5µm =10^4V/cm

1995 E_x=3.3.V/0.35µm=9.4 x 10^4V/cm

2001 E_x=1.2V/0.1µm=1.2 x 10^5 V/cm

Velocity Saturation

- As the horizontal E field grows, so does the velocity of the electrons
- As the gate shrinks, the velocity reaches the point where it cannot increase as much with increasing E field
- This is due to velocity saturation
  - Electrons become so energetic that they scatter (crash into the crystal lattice)

Alpha-Power MOSFET Model

T. Sakurai and A. R. Newton, "Alpha-power law MOSFET model and its application to CMOS inverter delay and other formulas," IEEE J. Solid-State Circuits, vol. 25, pp. 584–593, 1990.

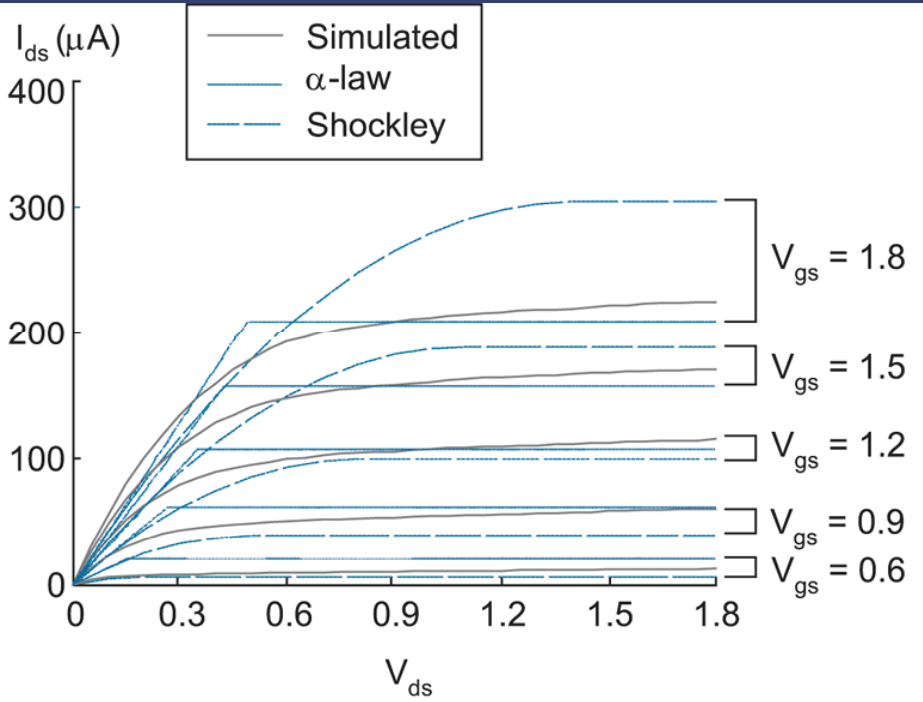


FIG 2.17 I-V characteristics for nMOS transistor with velocity saturation

At low lateral E-fields, V_{ds}/L, current increases linearly with E-field

At high fields, E= E_{sat}

Carrier velocity saturates due to carrier scattering = v_{sat} (= μ E_{sat})

$I_{ds} = \mu C_{ox} W/L (V_{gs} - V_t)^2$

---no velocity saturation

$I_{ds} = C_{ox} W (V_{gs} - V_t) v_{sat}$

---complete velocity saturation

Practical situation: carrier velocity doesn't increase linearly with field but is not completely velocity saturated….

Sakurai Model: $I_{ds} \propto (V_{gs} - V_t)^{\alpha}$

$1<\alpha<2$, is the velocity saturation index, determined by curve fitting…..also accounts for mobility degradation due to high vertical field (V_{gs}/t_{ox})

Mobility degradation

- MOS I/V equations depend on surface mobility μn (or μp)
  - Surface mobility is the mobility of the minority charge carriers at the surface of the channel
- In short-channel devices, μn and μp are not constant
  - As vertical electric field EY increases, surface mobility decreases because the electrons' wave function extends increasingly into the extremely low-mobility oxide
  - μ0 = low-field mobility, η is empirical constant
  - As VGS increases, effective surface mobility decreases

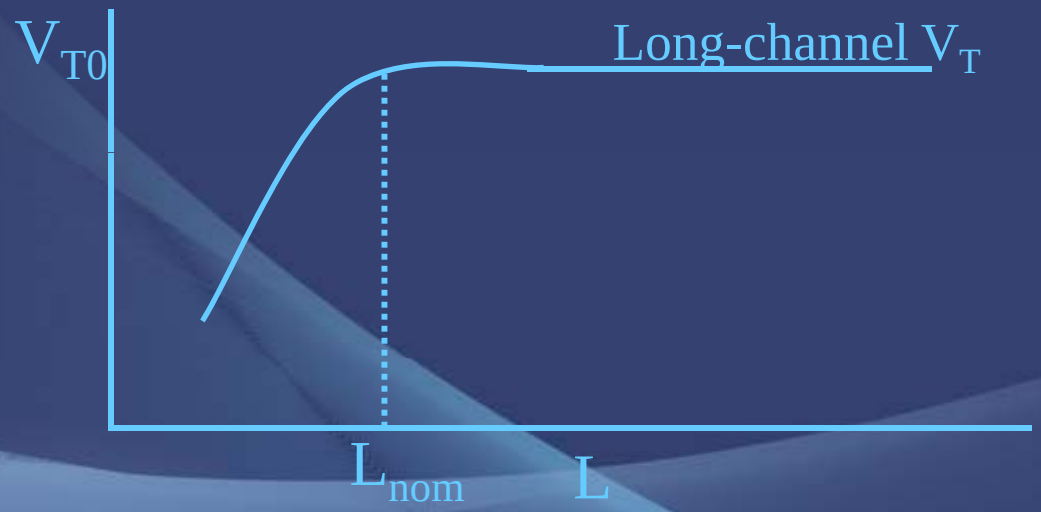
$$\mu_{eff} = \frac{\mu_0}{1+\eta(V_{GS}-V_T)}$$

Threshold voltage dependence on L

- Until now, threshold voltage assumed constant
  - V_T changed only by substrate bias V_{SB}
- In threshold voltage equations, channel depletion region assumed to be created by gate voltage only
- Depletion regions around source and drain neglected: valid if channel length is much larger than depletion region depths
- In short-channel devices, depletion regions from drain and source extend into channel

Threshold voltage roll-off…

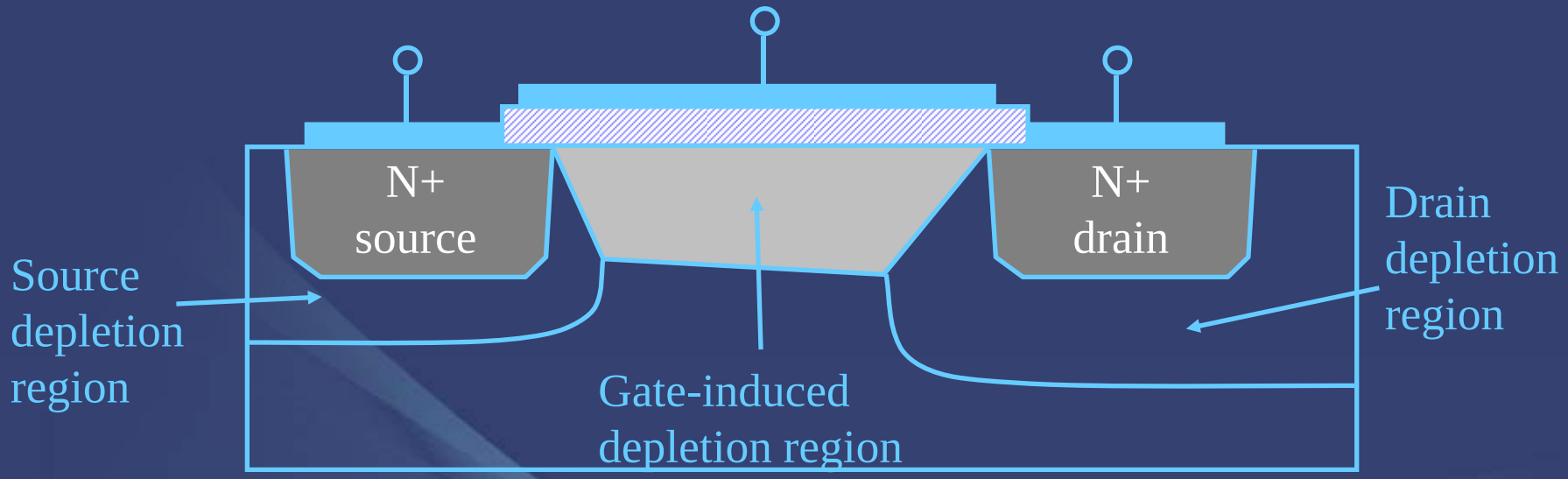
Graphically: V_{T0} versus channel length L



V_T Roll-off:

V_T decreases rapidly with channel length

Reason for Vt Roll-Off....



- Even with V_{GS}=0, part of channel is already depleted
- Bulk depletion charge is smaller in short-channel device $\rightarrow$ V_t is smaller

Threshold voltage roll-off….

- Change in V_{t0}:
  - x_{dS}, x_{dD}: depth of depletion regions at S, D
  - x_j: junction depth

$$\Delta V_{t0} = \frac{1}{C_{ox}}\sqrt{2q\varepsilon_{Si}N_A|2\phi_F|} \bullet \frac{x_j}{2L}\left[\left(\sqrt{1+\frac{2x_{dS}}{x_j}}-1\right)+\left(\sqrt{1+\frac{2x_{dD}}{x_j}}-1\right)\right]$$

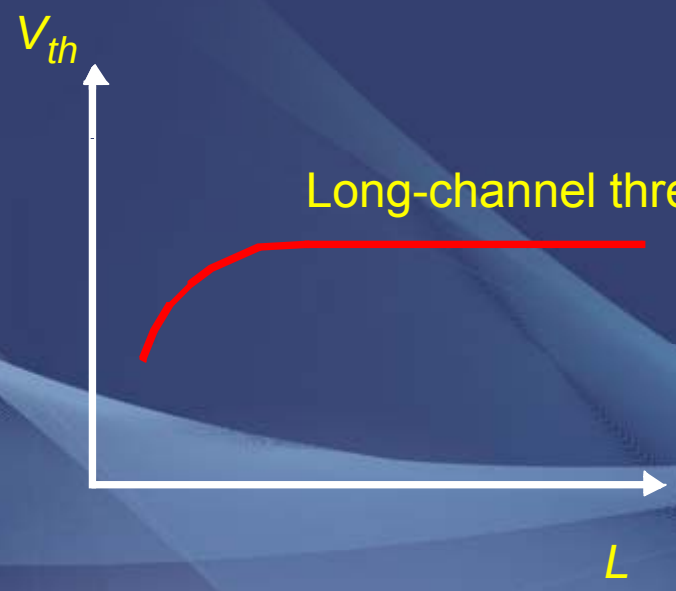
- ΔV_{t0} is proportional to (x_j/L)
  - For short channel lengths, ΔV_{t0} is large
  - For large channel lengths, term approaches 0

DIBL

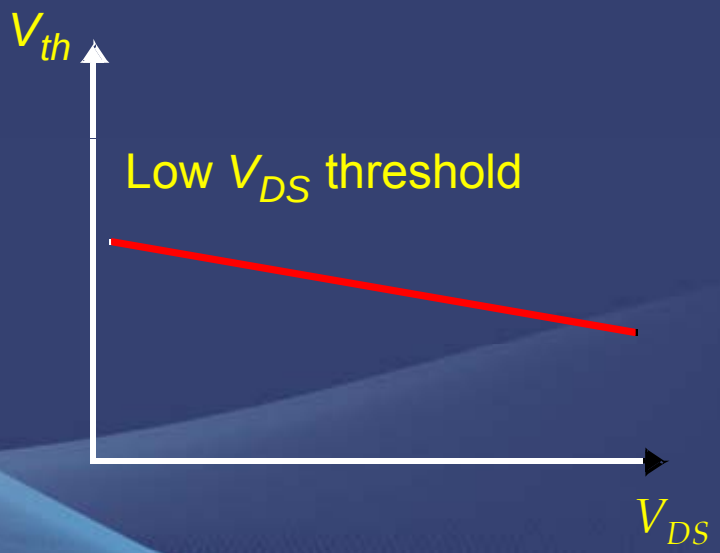
- Drain-induced barrier lowering (DIBL)
  - Drain voltage V_{DS} causes change in threshold voltage
  - As V_{DS} is increased, threshold voltage decreases
- Cause: depletion region around drain
  - Depletion region depth around drain depends on drain voltage
  - As V_{DS} is increased, drain depletion region gets deeper and extends further into channel
  - For very large V_{DS}, source and drain depletion regions can meet → *punch-through*!
- Issue: results in uncertainty in circuit design

Effect of SCE and DIBL on Vth…..

- $V_{th} = V_{th\_long\text{-}channel} - SCE - DIBL$



Threshold as a function of the length (for low V_{DS})



Drain-induced barrier lowering (DIBL) (for low L)

Threshold voltage variation

Short-channel effects cause threshold voltage variation:

- V_t rolloff
  - As channel length L decreases, threshold voltage decreases
- Drain-induced barrier lowering
  - As drain voltage V_{DS} increases, threshold voltage decreases
- Hot-carrier effect
  - Threshold voltages drift over time
- Negative-Bias Temperature Instability (NBTI)
  - Issue in PMOS transistors
  - Vt drifts over time
  - Typical stress temperature 100-150 C
  - Typical oxide electric fields of 5-6 MV/cm

Threshold voltage variation

- Hot-carrier effect
  - increased electric fields causes increased electron velocity
  - high-energy electrons can tunnel into gate oxide
  - This changes the threshold voltage (increases V_t for NMOS)
  - Can lead to long-term reliability problems

Threshold voltage variation

- Hot electrons
  - High-velocity electrons can also impact the drain, dislodging holes
  - Holes are swept towards negatively-charged substrate → cause substrate current
  - Called *impact ionization*
  - This is another factor which limits the process scaling → voltage must scale down as length scales

Threshold voltage variations

- Summary of threshold variations in short-channel devices
  - V_t rolloff: threshold voltage reduces as channel length L reduces
  - DIBL: threshold voltage reduces as V_{DS} increases
  - Hot-carrier effect: threshold voltage drifts over time as electrons tunnel into oxide
  - NBTI—causes Vt increase in PMOS transistors, strong dependence on Temperature.

Subthreshold Leakage

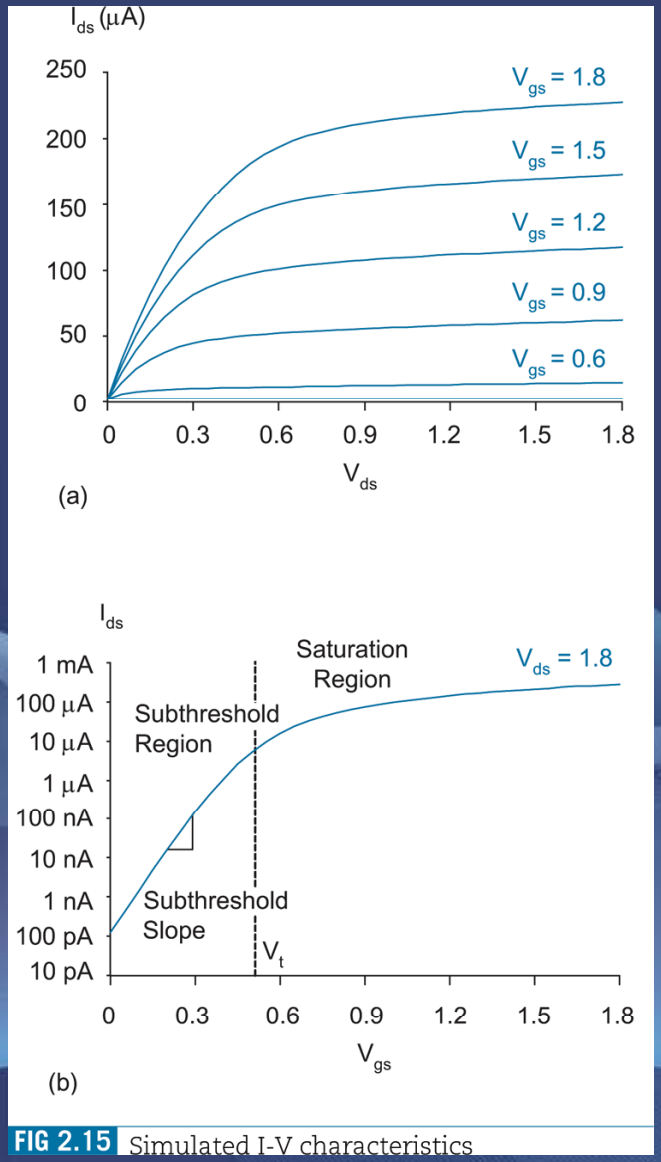


FIG 2.15 Simulated I-V characteristics

- *Dominant leakage mechanism*
- *Increases exponentially with temperature and Vt*

$$S = \left(\frac{d(\log I_d)}{dV_g}\right)^{-1} = \left(\frac{\partial V_g}{\partial \psi_s}\right)\left(\frac{\partial \psi_s}{\partial(\log I_d)}\right) = \left(1 + \frac{C_{dm}}{C_{ox}}\right)\frac{kT}{q}\ln(10)$$

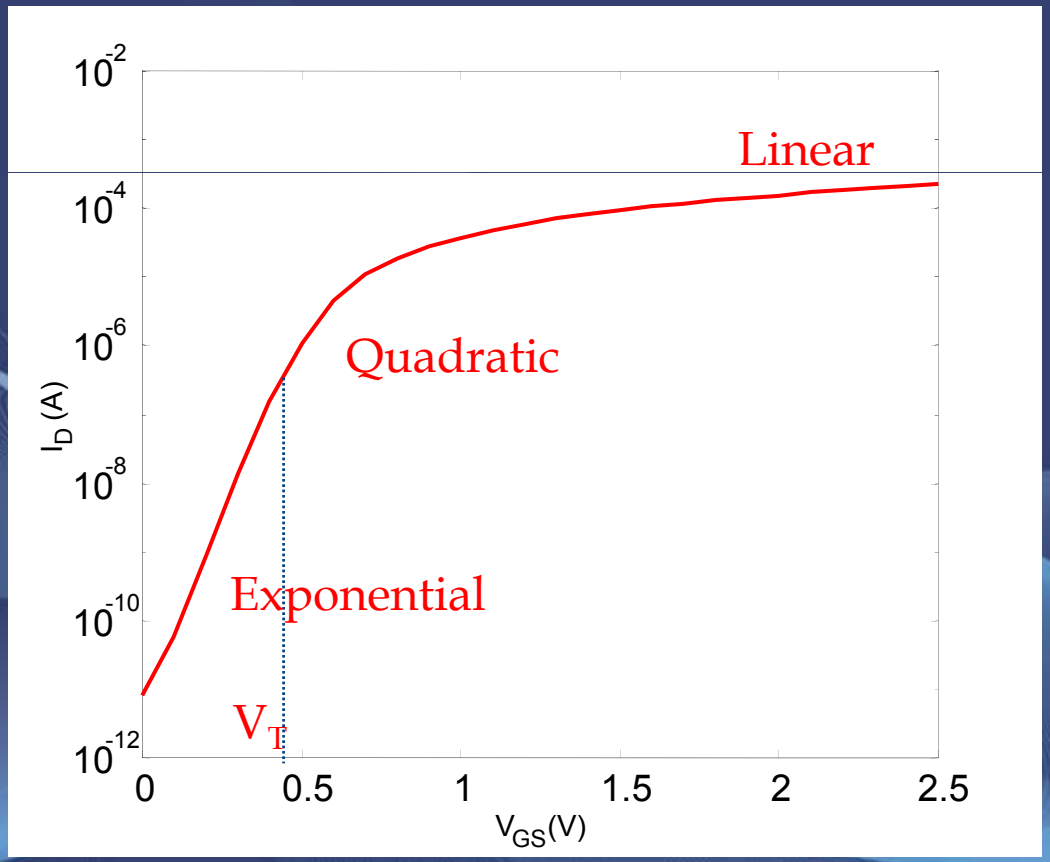
Sub-threshold conduction (1)

- When $V_{GS} < V_T$, transistor is "off"
  - However, small drain current I_D still flows
  - Called *subthreshold leakage* current
- Model for subthreshold current:

$$I_D(subthreshold) = I_S W e^{\frac{q}{kT}(AV_{GS} + BV_{DS})}$$

  - Increases as V_{GS} increases (potential barrier lowered)
  - Increases as V_{DS} increases (DIBL)

Sub-Threshold Conduction (2)



The Slope Factor

$$I_D \sim I_0 e^{\frac{qV_{GS}}{nkT}}, \quad n = 1 + \frac{C_D}{C_{ox}}$$

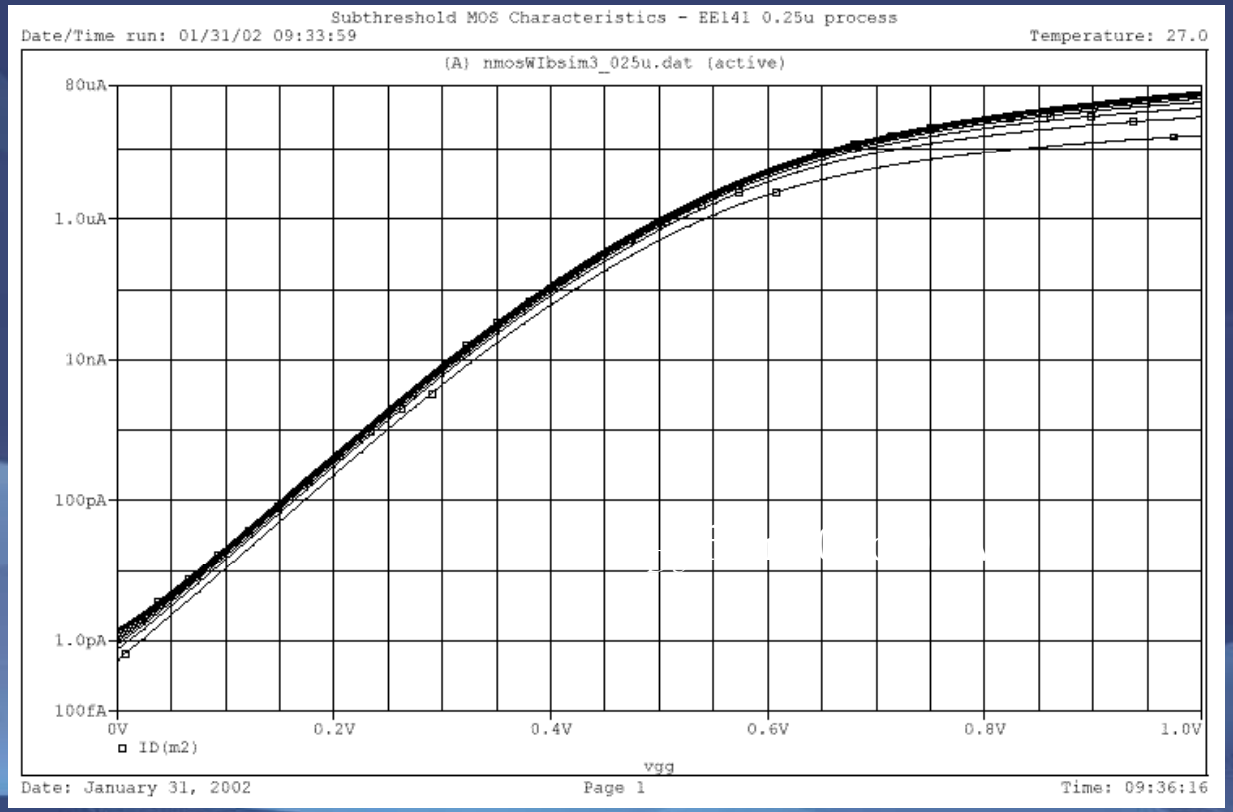
S is ΔV_{GS} for I_{D2}/I_{D1} =10

$$S = n\left(\frac{kT}{q}\right)\ln(10)$$

Typical values for S:
60 .. 100 mV/decade

Sub-Threshold I_D vs V_{GS}

$$I_D = I_0 e^{\frac{qV_{GS}}{nkT}} \left(1 - e^{-\frac{qV_{DS}}{kT}} \right)$$



Sub-Threshold I_D vs V_{DS}

$$I_D = I_0 e^{\frac{qV_{GS}}{nkT}} \left(1 - e^{-\frac{qV_{DS}}{kT}}\right)\left(1 + \lambda \cdot V_{DS}\right)$$